

Supplementary material for "Hack Weeks as a model for Data Science Education and Collaboration"

DANIELA HUPPENKOTHEN^{1,2,3,*}

—
ANTHONY ARENDT^{4,5}

—
DAVID W. HOGG^{3,2,6,7}

—
KARTHIK RAM⁸

—
JAKE VANDERPLAS⁵

—
ARIEL ROKEM⁵

¹*DIRAC Institute, Department of Astronomy, University of Washington, 3910 15th Ave NE, Seattle, WA 98195, USA*

²*Center for Data Science, New York University, 65 5th Avenue, 7th Floor, New York, NY 10003, USA*

³*Center for Cosmology and Particle Physics, New York University, 726 Broadway, 10th Floor, New York, NY 10003, USA*

⁴*Polar Science Center/Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Box 355640, Seattle, WA 98105-6698*

⁵*The University of Washington eScience Institute, The WRF Data Science Studio, Physics/Astronomy Tower, 6th Floor, 3910 15th Ave NE, Campus Box 351570, University of Washington, Seattle, WA 98105, USA*

⁶*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg*

⁷*Center for Computational Astrophysics, Flatiron Institute, 162 5th Ave, New York, NY 10010, USA*

⁸*Berkeley Institute for Data Science & Berkeley Initiative in Global Change Biology, University of California, Berkeley, Berkeley CA 94720*

Contents

1. Materials and Methods	2
2. Additional Statistical Results	3
3. Additional Figures	6
3.1. Career stage	6
3.2. Gender Identity	8
3.3. Racial/Ethnic Identity	10
4. Practical Considerations	12
4.1. What to do before	12
4.1.1. Obtaining permission for tracking and studying participants	12
4.1.2. Determining your objectives	12
4.1.3. Hack week Location	12
4.1.4. Event size and duration	13
4.1.5. Participant selection	14
4.1.6. Funding	16
4.1.7. Code of Conduct	18

* dhuppenk@uw.edu

4.1.8. Communicating with the outside world	18
4.2. During a hack week	20
4.2.1. Tutorials	20
4.2.2. Break-Out Sessions	22
4.2.3. Hack sessions	23
4.2.4. Facilitating participation	24
4.2.5. Impostor Syndrome	25
4.3. What to do after hack week	26

1. MATERIALS AND METHODS

We performed post-attendance surveys for AHW, GHW, and NHW in 2016 and 2017. All surveys contained general questions about attitudes towards the workshop, as well as open science and reproducibility, shared among all three surveys. The NHW and GHW surveys and the AHW 2017 survey were administered on site on the last day of the workshop; AHW participants in 2016 were e-mailed immediately after the workshop, with reminders after several weeks. All responses were collected within four weeks of the end of their respective workshops. The experimental procedures were approved by the Institutional Review Boards at UW, NYU, and UC Berkeley. All participants gave informed consent.

Response rates for NHW (2016: 41 responses; 2017: 45 responses) and GHW (2016: 42 responses; 2017: 41 responses) were 100 percent in both years; the response rate for AHW was 71 percent (35 out of 49) in 2016 and 82 percent (37 out of 45) in 2017. The lower response rates for AHW can be explained by the generally lower rates of response for the time-delayed survey in 2016. In 2017, a number of participants did not attend the full week, and thus several attendees were not present on the last day when the survey was administered. For the AHW surveys, we checked whether the demographics of the respondents differed significantly from those of the attendees overall in three relevant categories (career stage, racial/ethnic background, and gender identity) and found no significant difference for any of the groups except for career stage in 2017: Here, we find that graduate students are underrepresented (27 percent of respondents compared to 41 percent of attendees), while undergraduate students are overrepresented (25 percent of respondents compared to 12 percent of attendees). Because we group responses in the career stage category into the categories “early-career” (students and postdoctoral fellows), “late-career” (faculty, research scientists), and “other” (e.g. participants from outside academia), this discrepancy has no bearing on the results we report. For gender, we grouped participants into “minority” (women and non-binary) and “non-minority” (men) participants; similarly, for race and ethnicity, we divided participants into the categories “non-white” (all but Caucasian participants) and “white” (Caucasian). Note that demographic data for GHW and NHW is only available for 2016, while for AHW, data is available for both 2016 and 2017. Responding to any question in the survey was optional.

Participants were asked to respond to statements regarding these topics using a six-point Likert-type scale. All questions were anonymously recorded. No responses were discarded, and no pre-processing was performed on the data. We test for correlations between demographic characteristics (independent variables) and question responses (dependent variables) using a standard χ^2 test and compute the effect sizes via a bias-corrected version of Cramér’s V (1; 2), denoted ϕ_c in the manuscript. For a fairly permissive significant threshold of $\alpha < 0.05$, we require a p -value of $p < 0.0007$, corrected for $N_{\text{trials}} = 36$ trials, to claim a significant effect as well as at least a medium effect size according to (3).

We note that while a non-significant p -value can reject a hypothesis, this is not automatically evidence for the corresponding null hypothesis, since a lack of statistical power could also originate in a too small sample size. For cases where the null hypothesis corresponds to the scientific hypothesis of interest, we perform an equivalence test based on the Cramér’s V effect size following (4). Following a bootstrap procedure, we generate 10,000 random data sets from

the measured responses by sampling with replacement for each response variable and compute the corresponding effect size in order to generate confidence intervals around the measured Cramér’s V. For each hypothesis, we determine an equivalence bound for Cramér’s V for a medium-strength effect size based on (3) given the number of degrees of freedom and calculate the p -value that the measured effect size could extend above¹ this equivalence bound. If the latter hypothesis can be rejected, this indicates that the observed effect falls within the equivalence bounds and is statistically smaller than any effect deemed worthwhile.

The full analysis procedure is available online².

2. ADDITIONAL STATISTICAL RESULTS

Here we present the full results of the statistical analysis performed for the survey (Table 1). We tested for correlations between responses to a number of questions and demographic factors, most notably career stage, gender identity and racial/ethnic identity. The table also shows (1) the number of responses N to each question for each hack week, (2) value of the χ^2 and Cramér’s V statistics, χ^2 and ϕ_c , respectively, (3) the p -value associated with the χ^2 statistic, (4) the degrees of freedom for both χ^2 -test (dof_{χ^2}) and Cramér’s V (dof_{ϕ_c}), and (5) the upper 95% confidence interval for the Cramér’s V statistic (CI_U) as well as the corresponding p -value that the null hypothesis (the presence of an effect with moderate or higher strength) can be rejected, p_{eq} . The data sets used in this data analysis are presented in Section 3, Figures 1 through 12.

For no combination of questions with a demographic variable, do we find a p -value below the trial-corrected significance threshold of $p < 0.0007$ quoted in Section 1. In the main text, we place these null results in context, and refer to the tests that were near significance, and were coupled with a moderate to large effect size, as an indication of where future research may be fruitful. The results also discuss the present study’s limitations.

¹ Note that because Cramér’s V is defined on the interval $[0, 1]$, we only require an upper bound

² See the repository: <https://github.com/uwescience/HackWeek-Writeup>

Table 1.

Question	Astro Hack Week				GeoHackWeek				NeuroHackWeek															
	N	$\text{dof}_{\chi^2}^a$	p	$\text{dof}_{\phi_c}^b$	CI_{U}^c	p_{eq}	N	$\text{dof}_{\chi^2}^a$	p	$\text{dof}_{\phi_c}^b$	CI_{U}^c	p_{eq}	N	$\text{dof}_{\chi^2}^a$	p	$\text{dof}_{\phi_c}^b$	CI_{U}^c	p_{eq}						
Comparison with career stage^d																								
New topics, tools, methods (a)	61	10	5.64	0.85	1.93	0	0.26	0.12	36	6	2.51	0.87	1.88	0	0.31	0.16	41	10	15.53	0.15	1.90	0.23	0.32	0.17
Improvement of research (b)	61	6	9.89	0.13	1.93	0.18	0.24	0.08	35	4	12.41	0.02	1.88	0.35	0.28	0.13	40	6	7.42	0.28	1.90	0.13	0.31	0.14
Better science (c)	59	6	3.79	0.70	1.93	0	0.25	0.09	35	4	10.61	0.03	1.88	0.31	0.27	0.12	40	4	5.06	0.29	1.90	0.11	0.28	0.12
Teaching (d)	61	10	13.65	0.19	1.93	0.17	0.28	0.13	36	10	13.68	0.19	1.88	0.22	0.34	0.20	40	10	6.90	0.74	1.90	0	0.33	0.19
Gender Identity^e																								
Improvement of research (b)	60	3	2.08	0.56	0.98	0	0.26	0.02	35	2	5.26	0.07	0.97	0.30	0.33	0.07	40	3	7.62	0.05	0.97	0.34	0.30	0.05
Teaching (d)	61	5	7.30	0.20	0.98	0.19	0.30	0.05	36	5	12.71	0.26	0.97	0	0.46	0.11	40	5	3.17	0.67	0.97	0	0.36	0.11
Connection building (e)	61	4	3.70	0.45	0.98	0	0.28	0.03	36	3	5.89	0.12	0.97	0.28	0.34	0.08	41	3	4.42	0.22	0.98	0.18	0.32	0.07
Valued contributions (f)	57	4	5.98	0.20	0.98	0.18	0.28	0.03	34	5	9.46	0.09	0.97	0.36	0.34	0.08	38	5	6.75	0.34	0.97	0.21	0.35	0.10
Racial/ethnic Identity^f																								
Improvement of research (b)	59	3	0.97	0.81	0.98	0	0.27	0.03	34	2	0.01	0.99	0.97	0	0.34	0.08	39	3	2.79	0.43	0.97	0	0.32	0.06
Teaching (d)	61	5	5.63	0.34	0.98	0.09	0.31	0.05	34	4	1.29	0.86	0.97	0	0.39	0.12	39	5	3.06	0.69	0.97	0	0.38	0.12
Connection building (e)	61	4	1.64	0.80	0.98	0	0.28	0.04	34	3	8.42	0.04	0.97	0.40	0.36	0.09	40	3	3.49	0.32	0.97	0.10	0.33	0.07
Valued contributions (f)	58	4	13.14	0.01	0.98	0.40	0.29	0.04	32	5	2.49	0.78	0.96	0	0.40	0.14	37	5	4.05	0.54	0.97	0	0.36	0.11

Respondents to a post-event evaluation survey indicate their level of agreement on a six-point Likert-type scale to a range of statements (associated letters correspond to panels in Figure 2 in the main text). Results are for hack weeks held in 2016 and 2017 for AHW and in 2016 for GHW and NHW, respectively. The number of respondents for each question is included in the table above. Additional figures showing the distributions of responses for each question are available in Section 3 of the SM below.

^a While the range of Likert items was the same for every question, the degrees of freedom for χ^2 and Cramér's V may differ depending on the number of items with non-zero responses, and the number of non-zero items in the independent variable.

^b The degrees of freedom for Cramér's V corrected according to (2)

^c The bootstrapped upper 95% confidence intervals for the Cramér's V statistic derived according to (4).

^d Responses to the question of career stage were grouped into three categories: early-career stage; faculty and research scientists; other

^e Responses to the question of gender identity were grouped with respect to representation within the field of study into two categories: non-minority; minority

^f Responses to the question of racial and ethnic identity were grouped with respect to representation within the field of study into two categories: white; non-white

3. ADDITIONAL FIGURES

Here, we present a number of additional figures related to the survey data presented in the main text. The stacked bar charts presented here provide another view of the data that produced the results quoted in Table 1 above.

3.1. Career stage

Here we explore a number of survey responses with career stage as an independent variable.

"I hacked on topics, tools, or methods that were very new to me", career stage

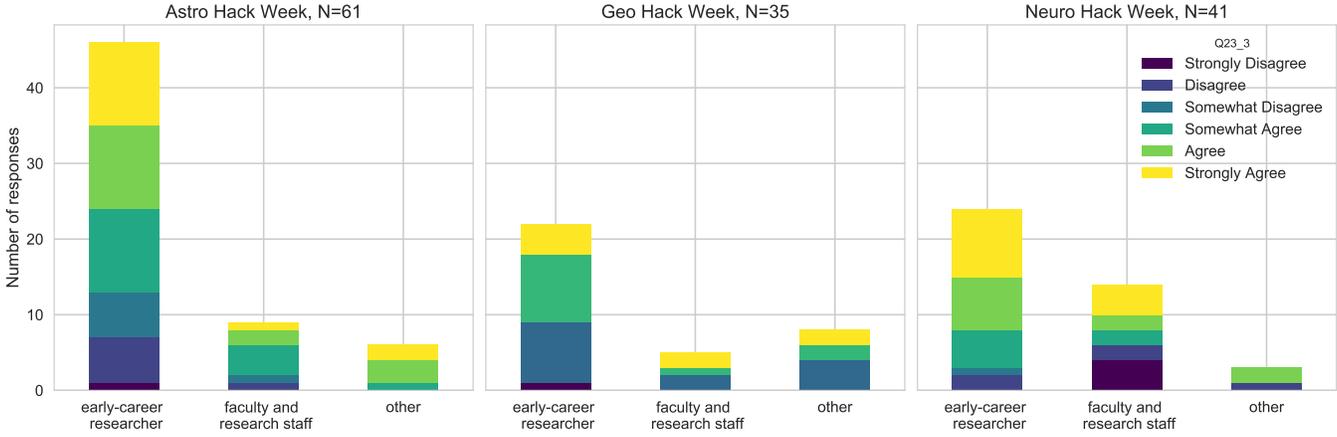


Figure 1.

"I believe that X Hack Week helped make me a better scientist", career stage

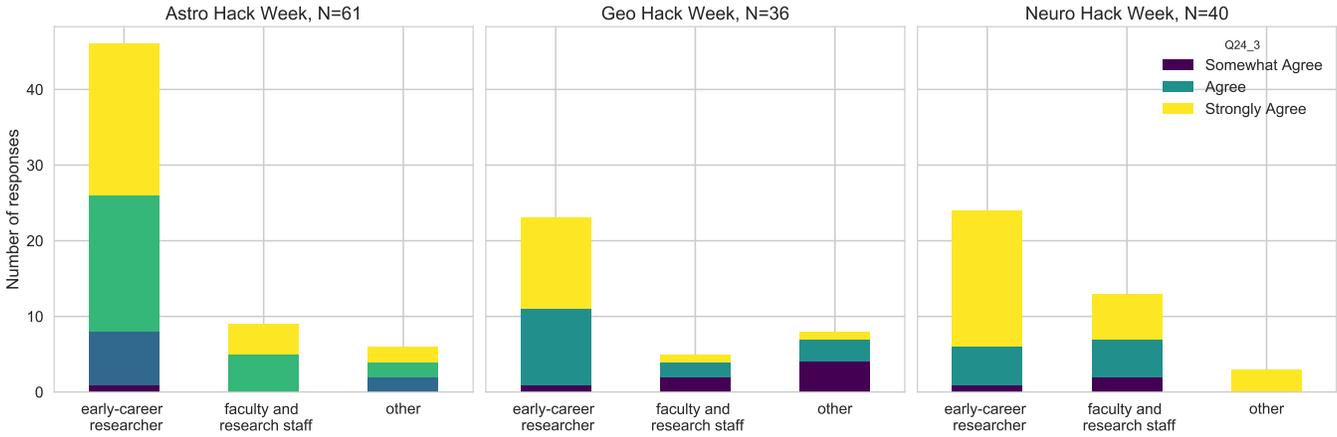


Figure 2.

"I feel like I learned things which improve my day-to-day research", career stage

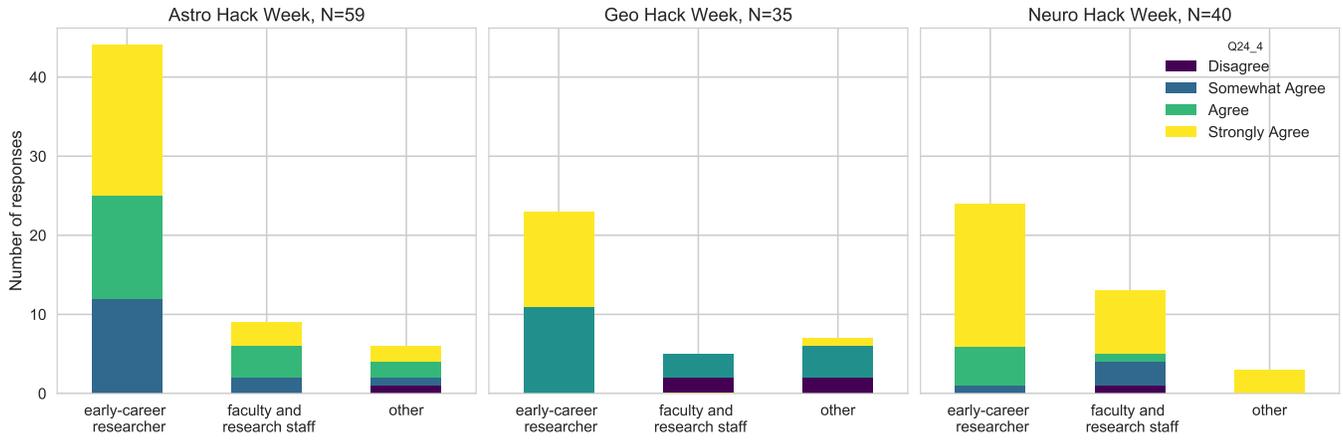


Figure 3.

"I ended up teaching things to other people", career stage

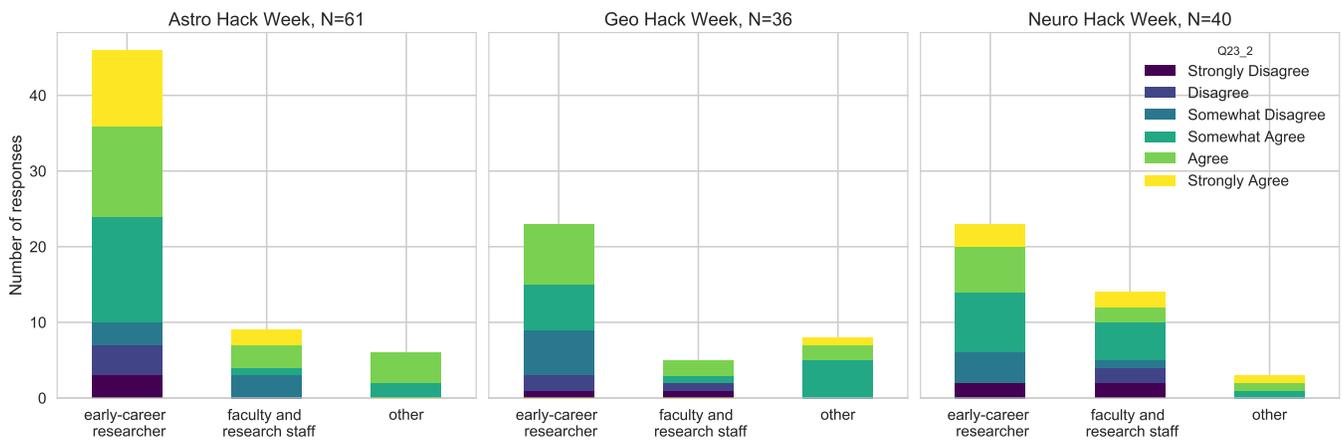


Figure 4.

3.2. Gender Identity

Here we explore a number of survey responses with gender identity as an independent variable. We grouped “gender identity” into two categories: “Non-minority” contains all participants who self-identified as male; “minority” contains all participants who responded with “female” or any other response. Women are here considered a minority with respect to their chosen field of study (neuroscience, astrophysics, or geoscience), where they tend to be under-represented.

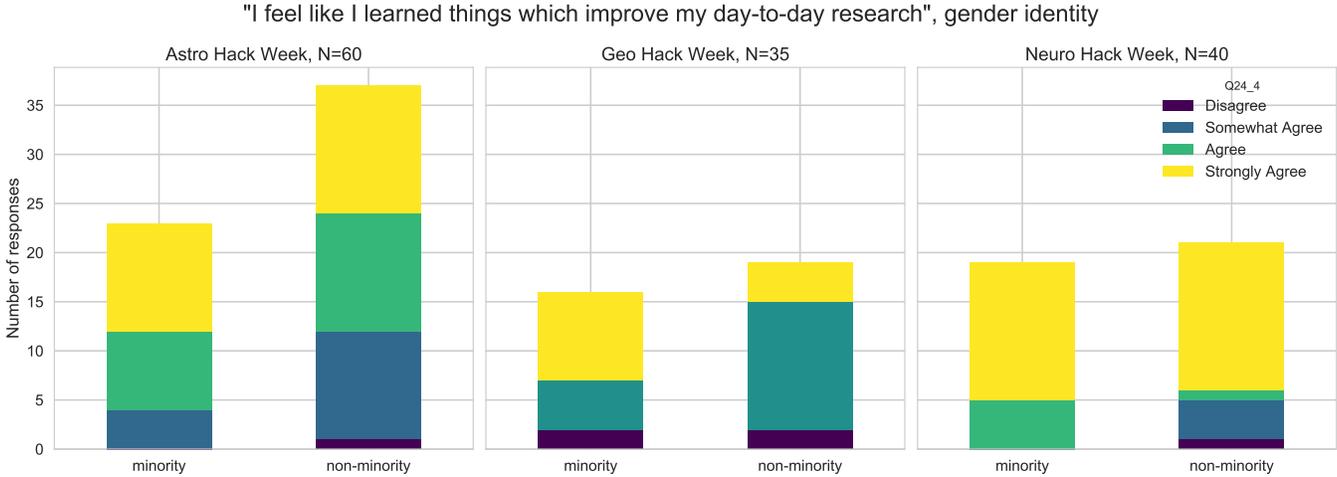


Figure 5.

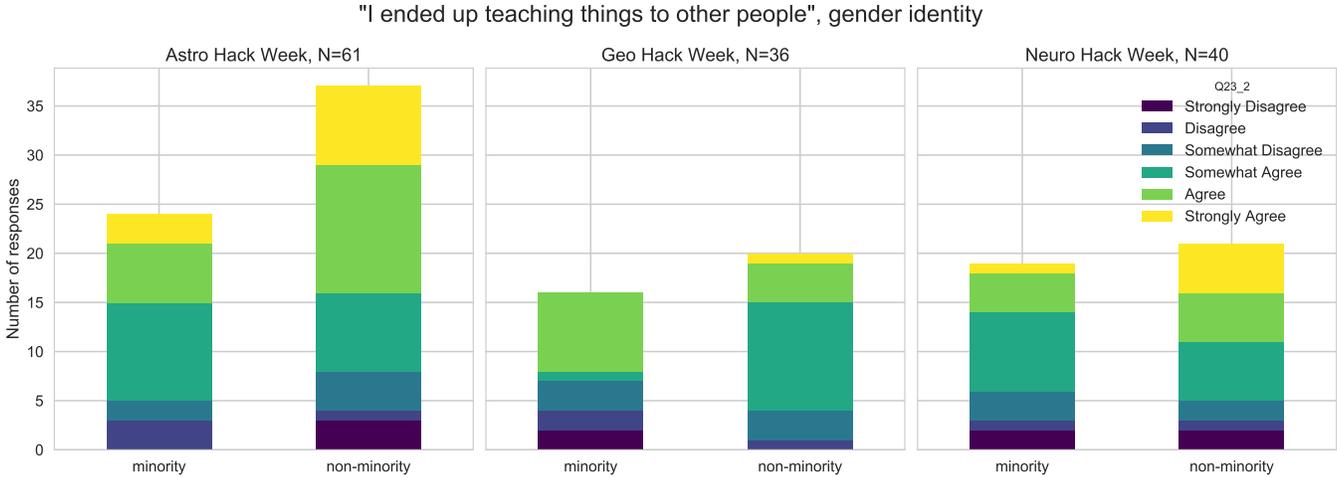


Figure 6.

"I felt like I built valuable connections at X Hack Week", gender identity

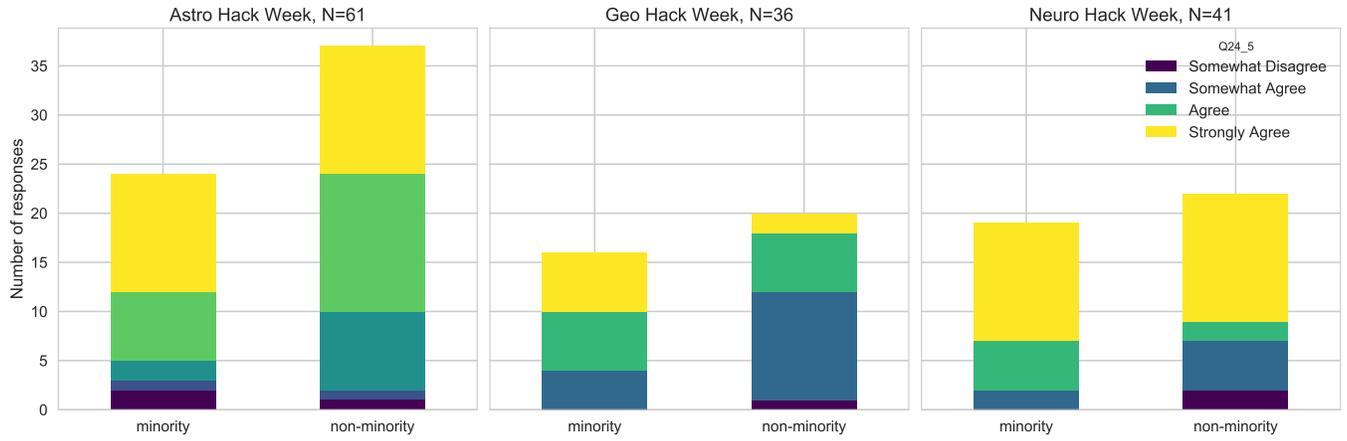


Figure 7.

"I felt like my contributions to my hack team(s) were valued", gender identity

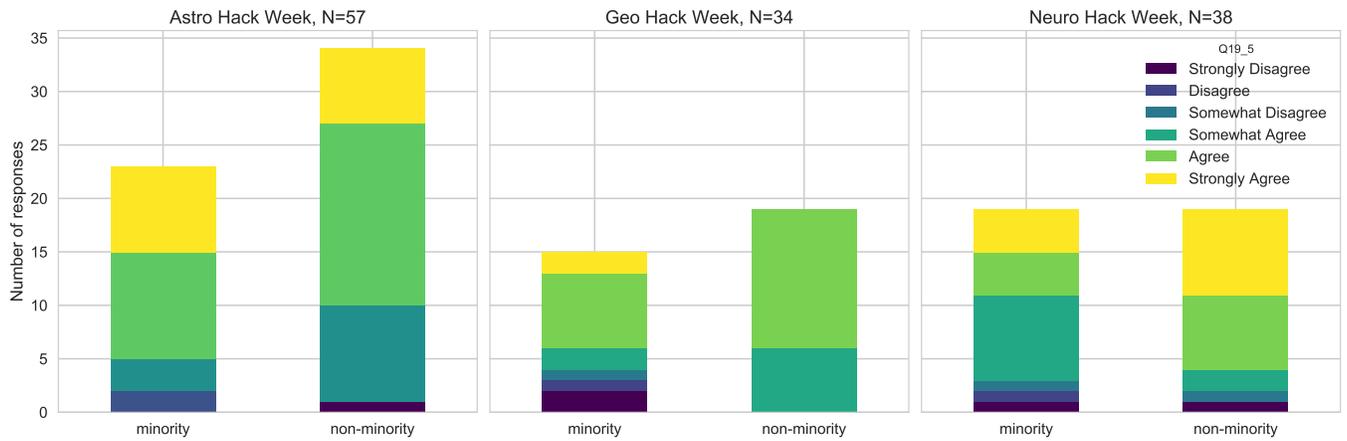


Figure 8.

3.3. Racial/Ethnic Identity

Here we explore a number of survey responses with racial/ethnic identity as an independent variable. We grouped responses into two final categories: “white” (Caucasian participants) and “non-white” (all others).

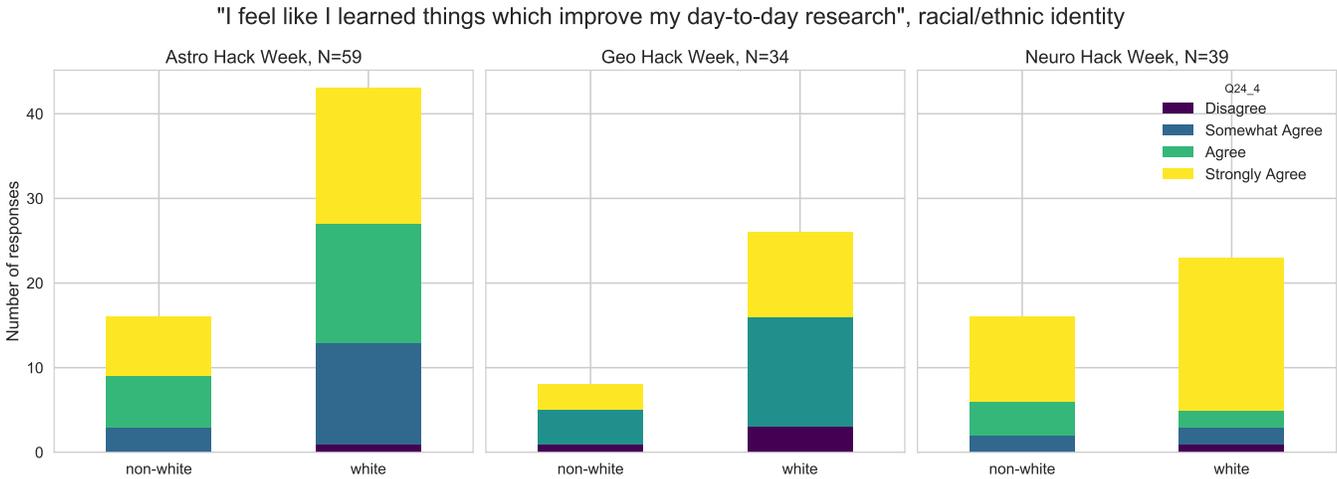


Figure 9.

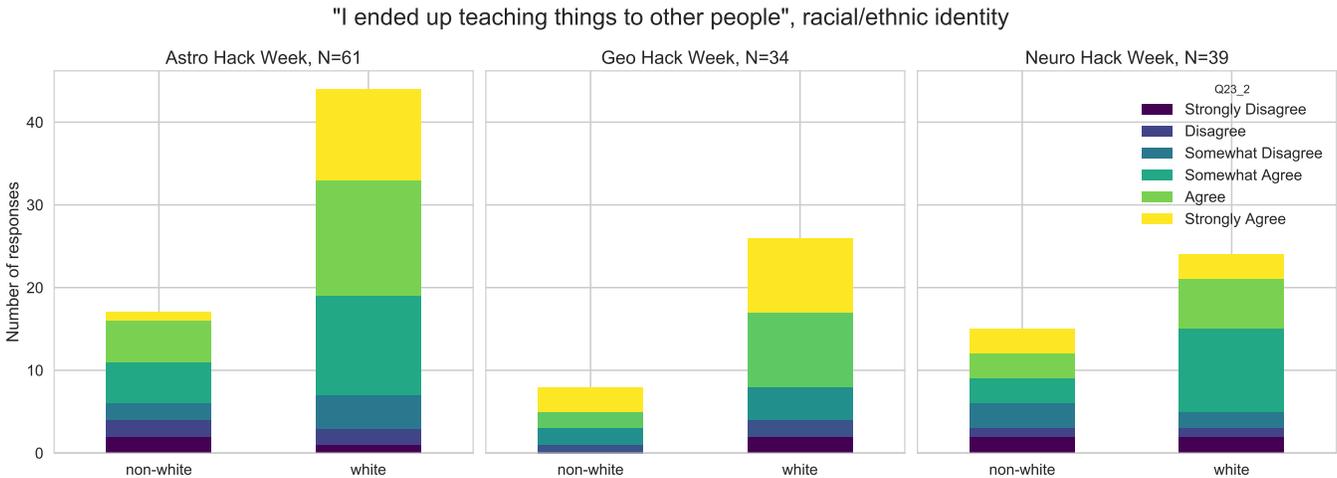


Figure 10.

"I felt like I built valuable connections at X Hack Week", racial/ethnic identity

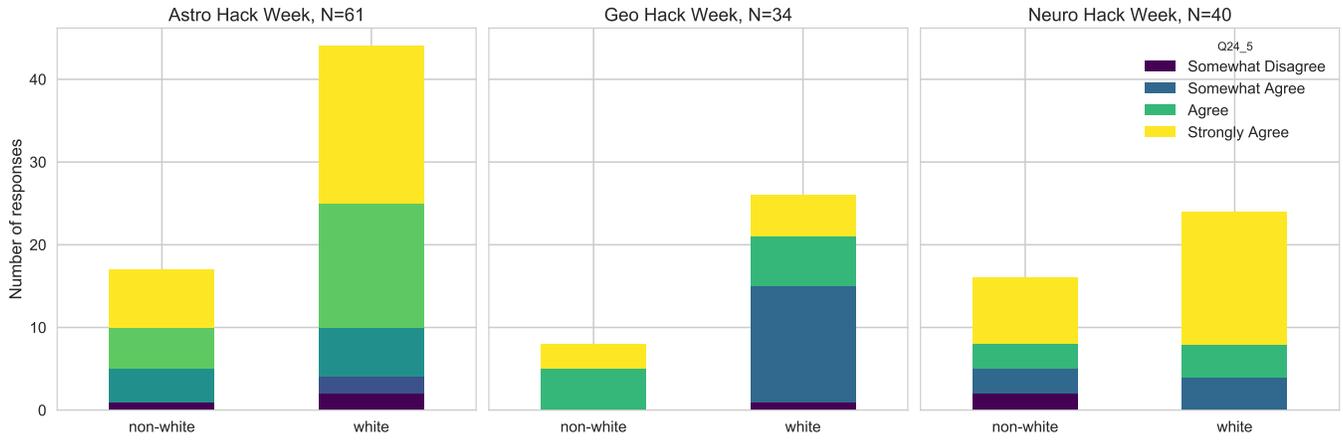


Figure 11.

"I felt like my contributions to my hack team(s) were valued", racial/ethnic identity

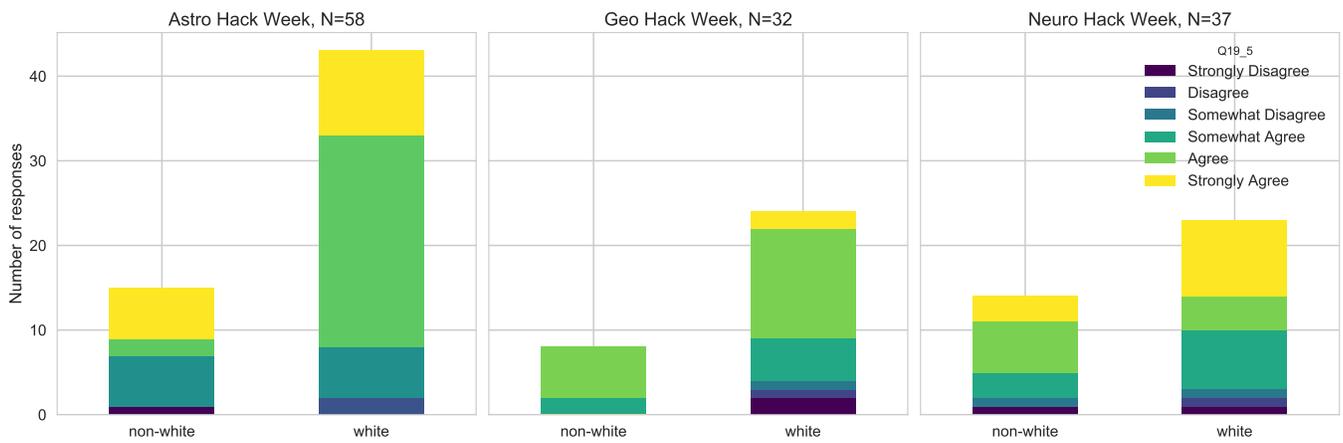


Figure 12.

4. PRACTICAL CONSIDERATIONS

4.1. *What to do before*

While the main text of this paper remains fairly theoretical, we aim to lay out more practical advice regarding the organization of a hack week in this supplement.

4.1.1. *Obtaining permission for tracking and studying participants*

A secondary goal of most of our previous events was to study the development of a hack week, to observe participant behavior, and to quantify the achievement of learning outcomes. To this end, we invited social scientists, interested in mechanisms of scientific collaboration and the educational design process to observe our activities. These researchers then reviewed their observations and gave valuable feedback about the hack week outcomes. In order to track and study participants, approval from an Internal Review Board is required. This should be done as early as possible, and having assistance from researchers in the social sciences is invaluable, in particular for hack weeks in fields that generally do not involve human subject research. Ideally, participants should consent to be contacted during the application stage, which opens up the potential for tracking both participants and non-participants and allows for evaluation of the hack week.

4.1.2. *Determining your objectives*

Unlike traditional conferences, where the program is primarily driven by speaker selection, careful goal planning is essential for a successful hack week. The stated goals will continue to determine almost all of the remaining organizational questions, including space, participant selection and workshop evaluation. Is the target audience made up primarily of students, or mature researchers, or a mix of both? Should the taught component dominate over project work or should it be only a minor program point? Is the community to be invited more software-focused or should outcomes skew towards publishable results? Does the hack week aim to maximize its output by carrying best practices into universities and institutions that may be particularly under-represented? These questions, among others, should have a very clear answer before any further organization takes place.

4.1.3. *Hack week Location*

When considering space, the broader location (city and country) and the specifics (the room in which the workshop will occur) should be taken into consideration. The non-traditional nature of hack weeks means that some students report challenges in receiving support from their supervisors or departments to attend. Thus, keeping the costs of attendance low (including travel costs, room and board, and conference fees) should be a high priority when choosing a location. Conference centers like the Lorentz Center at the University of Leiden provide funding along with the collaborative space, but have long lead times when applying for their spaces (typically 12-18 months). The same can be true for collaborative spaces within universities and active learning environments, which tend to be popular. Therefore, the search for a location should commence no later than twelve months before the workshop and ideally much earlier.

The space itself should be large and ideally configurable, with movable tables and chairs, several screens or projectors, and an ample number of white boards or chalkboards. Two useful guiding principles are to book a space that fits a group at least 25 percent larger than that envisioned for the workshop, and that the less configurable the space, the larger the extra room should be. There is a natural tension between keeping the group together and providing physically separated break-out spaces. On the one hand, conducting the workshop in a single large venue improves group cohesion and prevents self-segregation of researchers and ideas. On the other hand, the interactive nature of the workshop may lead to an environment in which it is difficult for individuals to focus on the highly complex tasks that are typical of hack week projects. Hence, it may be advisable to allow for diffusion of the group into adjacent rooms, while providing ample space in which the entire group can congregate for tutorials, breaks and reports out to the entire

group. There should be plenty of power outlets for participants to plug in their computers. The space should have ready access to Wi-Fi, for activities that require an Internet connection (e.g., collaborative software development), and organizers should arrange in advance to provide access for visitors from other institutions. Accessibility should be taken into consideration: Is the space accessible for participants who use wheel-chairs? Are accessible restrooms available?

4.1.4. *Event size and duration*

Event size (i.e. the number of attendees) and event duration are important to consider during the design and planning phase of a hack week. All hack weeks we have organized so far have had group sizes ranging from 40-60 participants. The number of attendees has depended on a number of different design considerations. One practical consideration was the size of the available location: The project-focused parts of a hack week, especially, generally require a reconfigurable space with more seats than attendees, such that spontaneous organization in sub-groups is easy. However, group size is also important in the context of community building, group dynamics and program decisions. We limited group size to a maximum of 70 participants, because previous anecdotal evidence has shown us that this is about the maximum group size in which interactions between all participants are still possible. We wanted to ensure that all participants had an opportunity to at least meet all other participants, in order to facilitate community building and keep the group from fracturing into cliques. Another important consideration is representation along the many axes of diversity we considered. For example, whereas a single undergraduate participant might feel like an outsider, admitting three or four can mitigate alienation. Similarly, in order to foster active collaboration among many different subfields, we endeavored to admit participants with many different types of overlapping expertise. Having too small a set of participants could interfere both with the goal of representation and that of fostering interdisciplinary collaborations and networking. Organizers of events for smaller groups (of less than ~ 30 participants or so) will need to attend to participant selection particularly carefully, in order to achieve the desired level of diversity across participants.

All hack weeks organized by us until now have lasted five days long. This duration was chosen to allow enough time for both learning and project work. Participants often arrived with ambitious project ideas, or with the goal of taking newly learned methods and applying them to their data sets right away. This meant that many projects tended to extend over multiple days. Attendees who participated in multiple projects often used the hack week to try out several different ideas. A five-day hack week allows for ample time for these ambitious projects or for experimentation, where a shorter hack week would necessarily have to be more constrained and focused. At the same time, hack weeks tend to be exhausting. Project work often extends into the evenings, and requires prolonged phases of high concentration of participants, thus we limited our hack weeks to five days in order to avoid excessive fatigue among our attendees.

It is worth noting, however, that these guidelines were developed out of design considerations rather than experimentation and testing. In the future, we plan to experiment with the format, duration and size of these events and continue to evaluate their interplay with our objectives.

4.1.5. *Participant selection*

As we lay out in the main text, participant selection is one of the most important design decisions made while planning a hack week. Conversations about participant selection, and planning of selection strategies should commence early, ideally at the start of the organizing process. Because selection procedures are subject to a whole range of human biases, it is worth thinking critically about established procedures and how to improve them. At the same time, hack weeks are comparatively lower-stakes events than other situations in which cohort selection is relevant, such as graduate school admissions. Thus, hack weeks may also serve as test beds for experimentation and improvements that can inform selection in other contexts.

The hack weeks presented here used different approaches, that nevertheless share several common themes and ideas. In particular, they all focused on making the selection procedure as quantitative as possible and use the data collected

during each stage of the process to critically evaluate the efficacy of the process and the degree to which it is subject to human bias. Secondly, much energy was devoted early on in the design of the application form, which often uses a combination of open-answer and multiple-choice questions to assess candidates' suitability, skill levels and backgrounds. Care must be taken when designing questions assessing skill levels, because options such as "beginner" and "expert" have no absolute scaling and may have different meanings to different applicants. They may also be subject to intrinsic biases, if candidates of certain backgrounds systematically underestimate or overestimate their skill levels compared to other candidates (the so-called Dunning Kruger effect). It is therefore advisable to calibrate skill levels using concrete milestones. For example, instead of asking participants whether they are "beginners" or "experts" at machine learning, we instead asked whether they had done machine learning for class work, had used techniques from machine learning in their own research, or had developed and implemented their own methods. Similarly, questions about programming skills may ask whether participants can perform a particular task in one of a list of languages (e.g., "write a module or a library"). Responding to questions about demographic characteristics should generally be optional, and organizers should consider how these questions might interact adversely with workshop goals, e.g. via stereotype threat. For example, Astro Hack Week only asked whether applicants considered themselves a minority with respect to gender identity within astronomy, rather than asking them to explicitly self-identify within a specific category. Meanwhile, Neuro Hack Week opted to ask about gender, ethnicity and race using the questions that are standard when reporting statistics about study participants.

For any kind of merit-based selection, committees should critically examine their assumptions about reasonable expectations regarding (minimum) skill levels required for participation. Candidates from underrepresented groups, smaller institutions or other countries may have had fewer opportunities to learn relevant skills. Consequently, in a merit-based selection based on these skills these questions may effectively trace demographic diversity and lead to biased selections even when committees are vigilant overall. If a strict merit-based selection is part of the procedure, it should be performed with these caveats in mind. It is always a good idea to check the results of any selection procedure for intrinsic biases. For example, comparisons of the group selected based on merit versus the group not selected with respect to various demographic variables may easily reveal unintended biases introduced at this stage, and the procedure can then be adjusted and refined.

Since human decision makers also routinely overestimate their ability to forecast future performance of candidates and underestimate intrinsic variance during selection procedures (5), it may be beneficial to set a fairly tolerant threshold for acceptance based on merit criteria, and select the cohort from this meritorious pool via an algorithm, imposing outside constraints on the selection based on the goals of the workshop.

One solution to the latter problem is implemented in the software *entrophy*³). The algorithm aims to find a group of participants that together match as closely as possible a requested distribution of specified dimensions (e.g., career stage, geographic location, etc.), to meet pre-set fractions set by the organizers. It is worth noting that this algorithm is vulnerable toward biases in two ways: First, humans will set the target fractions for any category of interest and any human biases involved in setting these target fractions will be perpetuated in the selection procedure. Second, perhaps more obviously, the algorithm can only act on information that has been collected. Biased participant sets may still result from selection procedures that fail to include crucial categories.

The overall selection procedure for Astro Hack Week was made up of a two-step process. In the initial step, we performed a blind assessment for suitability using open-ended questions querying candidates' motivations for attending the workshop, and their goals for attending. We only excluded a small fraction (< 10 percent) of applicants at this step, because our goals aligned with admitting a wide range of different skill levels and backgrounds. Applicants were generally either excluded for being an obvious mismatch with the workshop goals (as in the case of an Android

³ <http://github.com/dhuppenkothen/entrophy>

developer applying to work on an unrelated game) or for being duplicate entries. We subsequently used a range of categories queried with multiple-choice questions on the application to maximize diversity across these categories, which included career stage, gender identity, racial/ethnic identity, geographic origin, programming background, machine learning skills, statistics knowledge, and previous attendance at Astro Hack Week. Targets within each category were set according to the different workshop goals. For example, we set ambitious goals of admitting 50 percent attendees from under-represented minorities with respect to gender, race and ethnicity with the goal of increasing representation of these groups within astronomy and data science. Similarly, we set targets for other categories (e.g. skill levels) in order to maximize goals around collaboration, networking, and fostering an interdisciplinary exchange of knowledge. Subsequently, we used a greedy optimization algorithm to jointly optimize all categories simultaneously. We evaluated our results by directly comparing the characteristics of the selected set of participants with both our targets and the distribution of the set of all candidates. For demographic characteristics specifically, 42 percent of participants in both 2016 and 2017 considered themselves a minority with respect to gender identity in astronomy, and 32 percent of participants in 2016 (2017: 33 percent) considered themselves a minority with respect to racial or ethnic identity. About half of the attendees were students (~ 40 percent graduate students, ~ 10 percent undergraduate students), another 25 percent were postdoctoral fellows, and the remainder were about equally split between research staff, tenured and tenure-track faculty, and participants in non-academic careers.

Neuro Hack Week took a different approach: Participants provided a CV as part of their application, in addition to answering questions about their technical and research background, and writing a short statement of the purpose of their participation. Each application was reviewed by several instructors using 1-10 scales to rate several criteria: technical and computing experience, research experience, the quality of the statement, an assessment of the contribution of participation to the applicant's career, and an assessment of the degree to which participation would increase the representation of groups that are usually underrepresented in neuroscience and computing. The scores were combined, applicants were ranked according to the combined score, and the 40 highest-scoring applicants were then selected to participate. As an example, this procedure resulted in gender parity in 2016 and near-parity (17 women and 23 men) in 2017. In Neuro Hack Week, as well about half of the participants were students (though no undergraduate students participated), about 25 percent were post-docs, and the remainder included research staff and faculty members.

4.1.6. *Funding*

We have worked to minimize the cost of registration for our events in order to enable broad participation, especially for participants from institutions, countries and demographics that are generally underrepresented at other meetings. Locations that offer some funding or provide free space are preferable to those that do not. We have found that travel grants are of crucial importance especially to undergraduates, students from minority backgrounds and participants from underrepresented countries. We found that minority participants disproportionately decline to attend, and often cite lack of sufficient travel funding as the primary reason. In particular if increasing diversity in the more technical aspects of a field is one of the workshop's goals, providing enough travel funding for five to ten participants can make a significant difference in allowing minority researchers to attend. Funding for some activities can be obtained through commercial vendors. For example, cloud computing providers have educational programs that provide cloud computing credits for students. Other corporate sponsorships of certain activities or of the event as a whole can be used to lower participation costs, and make participation more feasible for many participants. Federal funding agencies are also sometimes able to fund hack weeks and similar activities under traditional funding mechanisms that support training. For example, Neuro Hack Week 2017 was supported through a grant from the National Institute of Mental Health through its standard R25 training grant mechanism, and this grant will continue to support it in extended format in the coming years.

4.1.7. *Code of Conduct*

An overarching goal of every hack week is to provide a welcoming, comfortable, and inclusive environment in which participants may learn and work. However, disparities in backgrounds, knowledge, and experience, together with the close collaboration that hacking requires, creates the potential for conduct that might offend participants, and preclude them from participating fully. For this reason, a hack week should have a clear and enforceable code of conduct, that can guide the behavior of participants, and help organizers maintain a professional and welcoming environment. It is the responsibility of the organizing committee to discuss what belongs in the code of conduct, consider how it should be enforced, and prepare for different scenarios and their possible resolution in advance. A part of this could and should be the recognition that participants (and, indeed, members of the organizing committee) may occasionally violate the code of conduct in minor ways, and the setting up of tools and mechanisms to resolve these infractions. Major infractions, conversely, should be taken seriously and in severe cases lead to exclusion of participants. Aside from important provisions banning discrimination on the basis of demographic characteristics, there may be domain- or workshop-specific clauses that may require recognition. For example, as communities move between programming languages or techniques, shaming participants for their choice of method or language should be strongly discouraged, since it will make participants feel unwelcome and defensive, and will inhibit learning on a larger scale. Even if no code of conduct infraction is expected, the very design of such a code sets expectations, and helps signal to participants that organizers have thought about creating a welcoming environment, which in turn feeds back to the creation of such a space.

4.1.8. *Communicating with the outside world*

Because hack weeks have enjoyed considerable popularity, they have often been oversubscribed. This leaves organizers with an important decision: Given the necessity of turning a large fraction of applicants away, what methods does the group wish to employ in order to communicate during the hack week with the wider community not in attendance? Tutorials may be live-streamed via popular services such as YouTube, but aside from the technical complexity of setting up a reliable live stream, their interactive format may not easily lend itself to the one-directional communication of a video feed. Even so, live streams and video recording posted after the fact have proven to be both useful and popular in the past, even while they require considerable effort to set up. At the very least, there should be some method of real-time communication with the in-group. Twitter has proven itself to be an effective medium of general communication with the wider community, including hack pitches, questions during tutorials and collaboration with researchers outside Astro Hack Week. However, the closed nature of Twitter to those with an account may not make it the ideal platform for participation. Other open-source formats that do not require sign-up, such as Matrix, may be more compatible with the hack week format ⁴. Important documents such as program descriptions, program changes, and hack descriptions should be open to the public and communicated clearly to facilitate collaborations with outside researchers.

It is, in principle, possible to organize an entire hack week virtually. However, the format rests firmly on a foundation of community-building and creating opportunities for participants to exchange ideas and knowledge. While these goals are not impossible to achieve in a virtual setting, both would be significantly harder to accomplish online than in a single physical location, to a point where the lack of immediate exchange may severely limit the generation of successful new collaborations and ideas. We thus recommend against running a purely virtual hack week.

⁴ <https://dave.cheney.net/2017/04/11/why-slack-is-inappropriate-for-open-source-communications>

4.2. During a hack week

Unlike a traditional conference, where the program is generally set by pre-selected talks, hack weeks require additional involvement from the organization committee during the hack week itself.

4.2.1. Tutorials

Tutorials provide a mechanism for rapid knowledge sharing and are often one of the most useful components of a hack week. Unlike traditional lectures, in which more time is available to explore numerous topics in depth, tutorials aim to deliver only the most essential information while ensuring that students remain engaged and interactive. Moreover, the need for an instructor to balance prepared material with spontaneous, interactive learning, while attending to community building and inclusive practices, can be difficult for even the most seasoned educator. Therefore careful planning and preparation is necessary for effective delivery of a hack week tutorial. Tutorials may also be video-taped and shared with a broader community. We have found that video recordings of tutorials are rather popular, with some of them drawing more than 1,000 views.

We have experimented with numerous designs in delivering our hack week tutorials in a way that is student-centered and focused on meeting all students at their particular skill levels. We have observed that, especially during long tutorials and with a particularly diverse group of students, there is a natural separation between participants who are focused on gaining new knowledge as it is being presented, versus those who already possess this knowledge and hence turn their attention to other subjects as the tutorial is underway. In practice, this concern can be mitigated by explicitly involving experienced participants in the tutorials. We encourage instructors to call upon experts in the audience to identify themselves and effectively act as teaching assistants. Expert participants often find this gratifying because it allows them to test their own understanding and improve their skills by learning through teaching. At the same time, beginners benefit by having increased interactions with other hack week participants, thereby increasing team cohesion and building community. Another technique for engaging expert participants during the tutorial is to invite them to work through more challenging content located in the online lesson, which will not be covered in the session, but for which they can obtain help during the tutorial.

Some of our hack week tutorials follow the structure of the Software Carpentry model (6). The content is divided into a series of lessons, each having a well-defined set of learning objectives, key questions, and expected outcomes. Interspersed throughout each lesson are a series of exercises that participants work through at their own pace. We frequently use a method drawn from Software Carpentry, whereby students indicate their need for help with an exercise via posting a colored sticky note on their computer. This provides a simple method for identifying those students who need help without interrupting the flow of the tutorial. We encourage instructors to develop online material to accompany their tutorials, either in the form of Jupyter Notebooks (7) or by following the markdown templates provided by Software Carpentry. This way, the hack week “curriculum” can be built up over time and students can refer to online content after the tutorial if they need to revisit certain topics. Depending on the content, some instructors convey concepts during the tutorial via “live-coding” while others display static code and their results. Either way, we find that dividing code into short blocks, each with a specific task, helps keep participants not to become overwhelmed with information. In some cases, tutorials require substantial software installation and data. Though we encourage participants to install these software dependencies on their own computers, so that they can continue using them after the tutorial, this is sometimes so complicated that it slows down teaching. In some cases, we have experimented with using *Jupyterhub*⁵ to provide students with immediate access to a computational environment that provides teaching materials (e.g., Jupyter notebooks), software dependencies, and data (8). Providing access to an ephemeral consistent environment for the purpose of teaching has become even simpler with the recent advent of the *Binder* system (9; 10).

⁵ <https://jupyterhub.readthedocs.io/en/latest/>

While the short duration of the tutorial format helps us minimize the problem of information overload, it can be challenging to decide on the scope and breadth of material to be covered. In-depth collaboration among instructors in advance of the event is particularly valuable in this regard, to assess which methods or subjects will be most useful for the specific audience. We suggest that the primary goals of the tutorials should be to provide an entry point into an exploration of participants' datasets, opening the door to more thorough study outside the hack week. Given the broad span of available tools and topics, the organizing committee should seek to teach content that represents the state-of-the-art and is deemed to be of greatest use to the broadest range of participants. A common theme across all our hack weeks is teaching initial tutorials in version control, command line interfaces, data science platforms and practices of reproducible research. These tutorials endeavor to provide a common baseline of techniques that participants can build on in other tutorials and project work. Subsequent tutorials are then delivered in more domain-specific fields, but can be arranged in a way that builds knowledge constructively over the course of the week. In some cases, it may be advisable to recommend some theoretical texts as preparatory reading for hack week, though speakers should not count on every participant arriving with the same knowledge baseline. Because of their practical nature, however, software requirements should be announced as early and clearly as possible, with the expectation that participants will have installed necessary software in advance. It may be helpful to designate a period of time early in the meeting to trouble-shoot installations before tutorials begin in order to avoid losing significant time during the actual tutorials.

It is worth noting that in practice, the mixed audience and interactive setting can often lead to impostor syndrome (see section 4.2.5 below) among the speakers, who lose their explicit status as expert among the participants. It is therefore imperative that speakers are made aware as early as possible about the audience, so that they can plan their tutorial accordingly. It also helps to make them aware that participants may be taking the role of teaching assistants during exercises, and encourage them to include participants in the teaching, rather than seeing them as adversaries who may seek to correct them. Conversely, experts may ask pointed questions about fundamental, important concepts, both in order to help the audience gain a deeper understanding and to encourage questions, especially from junior participants, who may be hesitant to ask questions due to their own impostor syndrome. At Astro Hack Week, we encourage members of the organizing committee and experienced participants to ask questions, particularly when they know the answer, but think the concept may not yet be clear to parts of the audience. This has in the past led to exchanges between experts that have greatly contributed to the tutorials in ways that were both unplanned and unexpected for all involved, beginners and experts alike. Including experts in the teaching also takes pressure off teachers during exercises, when demands on the speaker are generally high.

To summarize, a good tutorial will:

- Be very clearly tailored to the audience and narrowly scoped,
- Strictly limit the amount of lecture-style teaching to less than 50 percent,
- Use experts in the audience to ask key questions and act as teaching assistants during exercises, and
- Communicate technical requirements at least a week before the beginning of the hack week.

4.2.2. *Break-Out Sessions*

Because instructors and organizing committees are unlikely to know in advance what projects and data sets participants will bring, often there will be topics and methods not covered in the pre-planned tutorials that are of interest to the audience. Here, break-out sessions offer an alternative: short (30-45 minute) tutorials that are fairly spontaneously organized (with as little lead time as a few hours) and often taught by expert participants in the audience. These tutorials can be a more in-depth treatment of one specific method of interest (e.g. Gaussian processes, K-means clustering, or deep learning), or cover a practical skill that may be useful to the audience, but is not formally part of the

hack week (this is especially the case for skills related to software development, such as code testing, documentation, and profiling). Break-out sessions are generally more informal than tutorials and can be taught to a subset of the group rather than requiring all participants to attend. They should be limited to one or two a day, in order to allow participants to attend sessions without having to sacrifice a significant fraction of their time reserved for project work. It is also possible to intentionally leave a tutorial slot free, to be filled with one or two self-organized break-out sessions instead of a pre-planned tutorial. In practical terms, it is advisable to keep a physical board with requests and potential teachers in a prominent location during the week. Decisions on which option to choose from the list can be done via an informal voting process, under the condition that a volunteer to hold the tutorial has been found. Giving a break-out session can be a daunting task: While some experts may have relevant tutorials already prepared from other workshops, often participants are being expected to offer a thirty-minute talk with little to no preparation. Organizers can and should take steps to provide a positive, encouraging environment in which participants can volunteer their knowledge. They should particularly encourage junior participants, who may often be the most knowledgeable about the topics usually requested, to volunteer for break-out sessions. They provide a valuable teaching experience in a friendly environment and offer the opportunity to network with the larger community.

To summarize, break-out sessions:

- Consist of short tutorials about a specific topic or practical skill not covered in the tutorials
- Are more informal and may be taught to a subset of interested participants
- Should be limited to one to two per day and take no longer than 45 minutes to avoid conflicts with project-work
- Can be daunting to teach, especially for junior participants, and
- Require the organizing committee to be pro-active about encouraging volunteer teachers.

4.2.3. *Hack sessions*

For hack weeks that place less emphasis on teaching, hack sessions are at the core of the workshop. They provide the opportunity for participants to actively collaborate and in many cases use the knowledge and skills learned in the tutorials on their own data sets. Even though hack sessions are strongly driven by the participants themselves, they require careful planning and vigilant supervision by the organizing committee. In general, hack sessions follow a standard pattern. In the initial stage, participants pitch projects to the group and in some cases request specific expertise (of a certain method, programming language or other skill). Ideally, the organizing committee will encourage participants to post potential projects as early as possible online in a central document, which will allow some organization even before the beginning of the hack week. If the hack week admits very junior participants (e.g. undergraduate students or junior graduate students) it may be advisable for the organizing committee to contact experienced participants in advance and ask them to suggest hacks that are appropriate for these junior participants. At the hack week itself, the time designated for project work will generally be prefaced by a pitch session, where participants with projects give a short description of their project to the group and request help. For junior participants and those new to the format, this can be daunting, even if (or especially when) they do not have a project themselves. Junior participants, in particular, tend to believe that they do not yet have the skills to contribute meaningfully to any projects. Here, it can be useful to require the participants who are not pitching projects to state their expertise and affirm that every participant has expertise in something. This broad definition of hacking can help the organizers to facilitate inclusion by emphasizing the value in skills such as writing (e.g. for tutorials and documentation), testing code to be released, or other tasks. After pitching, participants self-organize into teams to work on specific projects. Here, organizers should remain aware of participants who may not have found a group, or are having trouble finding a task to work on. This remains true

during the week, as some projects are abandoned and others appear. At least one member of the organizing committee should be familiar with the current roster of projects and the experience in the room to help match participants as needed. Tracking ongoing projects can be accomplished online, via real-time collaborative documents, as well as in the physical space, via a diagram of where in the space teams are located. Because hack weeks are generally longer than typical hackathons (which are often limited to just one or two days), this will lead to a variety of shorter projects completed within a day or two, and longer projects going on throughout the week. Regular check-in sessions are helpful for pitching new projects, requesting help on projects that are stuck at a certain point, and re-matching participants after hacks have been completed or abandoned. These sessions can also be used for completed hacks to be showcased, or intermediate achievements to be presented. At the end of the week we provide ample time for participants to present their projects to the community. We endeavor to design these presentations in a way that builds confidence, especially for junior researchers, and that welcomes all projects regardless of their level of complexity and/or sophistication. The open-ended format of a hack week encourages experimental approaches and new directions. This means that some projects will not reach a clear or successful conclusions. Too often, showcases focus on the most successful and impressive projects. This tends to distort the reality of the experience of hack week participants and produces both an unreasonable expectation of what the result of a hack week ought to be, as well as a pressure to be successful that can in practice inhibit participants. Therefore, it should be encouraged that participants show failed ideas, not least because the knowledge of why a project failed can be a valuable learning experience beyond the team working on that particular project.

4.2.4. *Facilitating participation*

Admitting a diverse group of participants is only the first step in encouraging collaboration and knowledge exchange. Community building and careful facilitation of both tutorials and project works are arguably the cornerstones of a successful hack week. This is especially important because the group will inevitably include a wide range of personalities, some of whom may find it easier to form new bonds and interact with the group than others. At all the hack weeks we have organized, we have taken concrete steps to address potential issues and foster a welcoming, inclusive atmosphere. Clearly communicating values and expectations at the beginning of the week can help set the stage for an environment in which a large range of different personalities can thrive. This includes both communication of the code of conduct, putting an emphasis on collaborative rather than competitive behavior, as well as a discussion of potential issues that participants might experience, such as impostor syndrome (see also section 4.2.5 below).

Our experience indicates that there are ways in which organizers can influence group dynamics in specific ways to allow timid participants to speak up. As mentioned above, we ask a few participants known to us to occasionally ask questions during tutorials, in order to encourage other participants to do the same. We also designate at least two members of the organizing committee to take on the role of facilitating both tutorials and hack sessions. It is their task to also manage dominant personalities, which can be both a boon and a hindrance to the success of a hack week. Often, these attendees can be effective at generating excitement and engagement, but can also intimidate others and dominate discussions. It is up to the organizers to balance the participation of these attendees with the needs of the rest of the group. Similarly, facilitators at the hack weeks we have organized spend the majority of hacking time circulating the room, connecting participants without a group to a project, helping connect groups who run into a problem with experts who can help out, and making sure that group work proceeds in such a way that all members feel valued.

At Geo Hack Week, we employed “Liberating Structures”, a facilitation approach in which we introduce small shifts in the way groups engage in order to increase relational coordination and trust. We invited a professional facilitator to assist us in leading a series of conversations and activities aimed at fostering group engagement. In one activity, participants faced one another and were given a brief statement inviting a two minute conversation, after which the rows rotated to maximize the mixing of ideas. Invitations included statements such as “I belong at Geohack Week because...”

or “A challenge I hope we can make progress on is...”. Importantly, event organizers joined in each activity, which we found helped participants feel as if they were not alone in facing some of the challenges inherent in data science studies. Another activity was designed to address the challenges of forming project teams. Individuals were invited to briefly pitch project ideas to the full group, and then form stations around the room. The remaining participants were then invited to visit any station that drew their interest, and to rotate between up to two other stations if they desired. We found that this approach minimized the pressure on participants, giving them time to get a feel for project ideas and group dynamics before joining a team.

4.2.5. *Impostor Syndrome*

The *impostor phenomenon*, or *impostor syndrome* (IS) is a dissonant feeling experienced by certain high-achieving individuals, that despite objective evidence to the contrary, they are in fact not as intelligent or capable as they appear. Individuals with IS thus experience a fear of being “found out”, shamed and expelled from their environment (11). Initial observations of the first Astro Hack Week conducted by data science ethnographer Brittany Fiore-Gartland⁶ suggested that hack weeks are an environment prone to a particular kind of IS, namely that participants feel the need to be experts in multiple scientific and technical aspects of the activities pursued during a hack week. This particular form of IS hinges to some degree on the design focus on diversity of backgrounds in that all the other participants seem to know something that any given participant does not. This might be further exacerbated by the expectation that attendants expose their ideas to public scrutiny, find collaborators in a very short amount of time, produce and present a successful hack at the end of the week⁷. The prevalence of IS at a hack week may be endemic to the format, and should thus be a major concern for any organizing committee. This is because of the chilling effect it tends to have on participants and the community as a whole, and particularly on women (in particular in fields in which women are under-represented) and members of ethnic and racial minorities, correlating with anxiety and other forms of mental distress (12). Less severely, another major concern is that IS inhibits risk taking: Participants experiencing it will be less likely to ask a question, to put forward an idea for a hack, and to be pro-active about forming new collaborations. Many of the goals of a hack week, including the successful completion of projects, lateral knowledge transfer, and community building are hampered by IS. Thus, we are working within our hack weeks to mitigate IS using various techniques. With respect to minority participants, ensuring adequate representation can decrease feelings of otherness and may help reduce IS. More generally, being open about the presence and prevalence of IS can help participants feel more at ease. Additionally, role models are very effective at encouraging positive behavior: Asking participants with prior experience at hack weeks at all academic levels to ask questions, even when they might know the answer, can help foster an inclusive environment that rewards risk taking. We also invited our organizing committees to model open communication about their own handling of IS through the course of their career. We found that transparency about one’s own uncertainty, especially from community leaders, helped create a hospitable atmosphere and removed barriers.

4.3. *What to do after hack week*

Feedback is a crucial component of a hack week. Because of its experimental design, and the differences in group from year to year, it is important to routinely check whether the workshop still matches up with what participants expect and find useful. Additionally, computational fields tend to be fast-paced; a tutorial given one year might be outdated a few years later. Therefore, it is useful to put some detailed thoughts into the survey. Following discussions the organizing committee will have likely had during the initial stage of preparing the announcement and selecting participants, organizers will want to confront their own expectations before the workshop, and design a survey that allows for both measurements of specific outcomes and goals the organizers are interested in, as well as leaving enough

⁶ <http://astrohackweek.org/blog/ethnographic-notes.html>

⁷ see also this insightful blog post: <https://medium.com/astronomy-without-stars/the-horror-of-hack-days-52c6b52cfc3b>

free space for participants to self-report outcomes that arose spontaneously or were outside the scope of the initial design. It is helpful to design the survey early and allow for some time near the end of the workshop for participants to fill it out while they are still physically present; in our experience, this leads to much higher response rates than if the survey is distributed after the end of the workshop. While some improvements and changes in survey design from year to year are inevitable, it is advisable to set a number of core questions that do *not* change from year to year. These questions will allow organizers to assess the workshop’s function within their community, and assess how changes to the format between one year and another may affect outcomes of interest.

For tracking tangible outcomes, a central real-time document used during the workshop to keep track of projects can function as a record of productivity. This document might be rewritten as conference proceedings. Additionally, organizers should encourage participants to identify written outcomes of the hack week: For publications, acknowledgments provide a convenient venue. Code repositories can allow this kind of acknowledgment in a *readme* document; GitHub also allows attaching tags to a repository, so all code repositories that experienced major contributions during the workshop may include a tag specific to the hack week.

On a larger scale, it might be useful to track measures of success that are more long-term. Those could for example be publication networks for scientists who attended the workshop, including their publications as well as their career development. For this purpose, it is useful to keep a way to contact previous participants, via a mailing list or real-time messaging services like *Matrix* or *Slack*.

REFERENCES

- [1]H. Cramér. *Mathematical Methods of Statistics*. (Almqvist & Wiksells Akademiska Handböcker). Princeton University Press, 1946.
- [2]Wicher Bergsma. A bias-correction for cramer’s v and tschuprow’s t. *Journal of the Korean Statistical Society*, 42(3):323 – 328, 2013.
- [3]Jacob Cohen. *Statistical power analysis for the behavioral sciences 2nd edn*. Routledge (New York), 1988.
- [4]T. Shiskina, L. Farmus, and R.A. Cribbie. Testing for a lack of relationship among categorical variables. *The Quantitative Methods for Psychology*, in press.
- [5]Scott Highhouse. Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, 1(3):333–342, 2008.
- [6]Greg Wilson. Software carpentry: Lessons learned. *F1000Research*, January 2016.
- [7]Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian E Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B Hamrick, Jason Grout, Sylvain Corlay, et al. Jupyter notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87–90, 2016.
- [8]C. Holdgraf, A. Culich, A. Rokem, F. Deniz, M. Alegro, and D. Ushizima. Portable learning environments for hands-on computational instruction: Using container-and cloud-based technology to teach data science. In *Practice and Experience in Advanced Research Computing, 2017*, 2017.
- [9]Chris Holdgraf. Binder 2.0, a tech guide – jupyter blog. <https://blog.jupyter.org/binder-2-0-a-tech-guide-2017-fd40515a3a84>, November 2017. Accessed: 2017-12-8.
- [10]C Titus Brown. Four steps in five minutes to deploy a carpentry lesson for a class of 30. <http://ivory.idyll.org/blog/2017-four-steps-five-minutes-binder.html>, December 2017. Accessed: 2017-12-8.
- [11]Pauline R Clance and Suzanne A Imes. The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention. *Psychotherapy: Theory, Research & Practice*, 15(3):241, 1978.
- [12]A Parkman. The imposter phenomenon in higher education: Incidence and impact. *Journal of Higher Education Theory and Practice*, 2016.